STYLEFORMER

A Convolution-Free Style Image Generator based on Transformer And Stylegan2.

Sapienza
University of Rome

Fabio Caputo

Weihao Peng

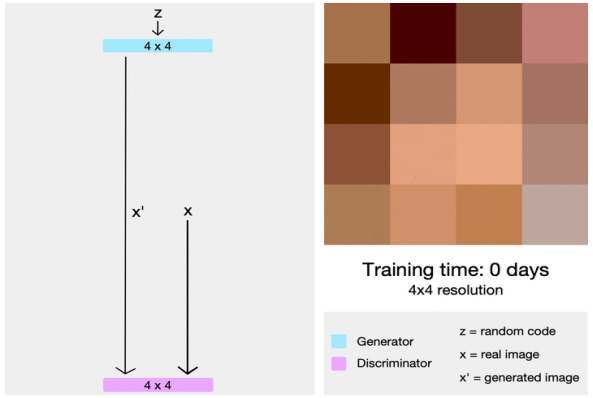
STRONG STYLE GENERATOR CONVOLUTION - FREE

GAN's (Generative Adversarial Networks) models are living a huge success since they were introduced in 2014, nowadays resolution and quality of the generated images increased a lot, what does not change is the consideration of convolutional operations as fundamental to achieve high-resolution images and a stable training. In this work, we have tried to implement a strong, but also light, style-based generator with a convolution-free structure, based on NPL technologies such as Transformer and Attention.

TECH OVERVIEW

How GAN works and why we are trying to use NPL technologies to generate images.

- Frist GAN model was introduced by Ian Goodfellow in 2014.
- StyleGAN is a progressive growing GAN architecture, able to synthesizing high resolution and quality images with incremental growing of discriminator and generator.



- That model shows some problems in generation, StyleGAN2 addressed most them using skip connection and replacing AdaIN with a statistic-based approach.
- Some problem remains, shortcomings derived using a convolutional network such locality problem led to a difficult capture of the global features.

STYLE GENERATIVE ADVERSARIAL NETWORK

NEURAL NETWORK 2022

ALL YOU NEED IS TRANSFORMER

- *"The first transduction model **relying entirely on self-attention** to compute representations of its input and output without using sequence-aligned RNNs or convolution"*
- Designed for NPL, recently is rising as an alternative to convolution operation in the computer vision field.
- Based on **attention**, a mechanism that mimic the cognitive attention focusing on small but significative details of an image, a token or any other significative data.
- Stacking attention and combining them with feed-forward layers, we can form encoders (**self-attention**).

ALL YOU NEED IS TRANSFORMER

- Solves the difficult to capture long-range dependency without stacking multiple layers.
- Indeed, using self-attention we are able to capture long-range dependency and understand global features efficiently.
- Using Linformer we can address Transformer expensive cost while dealing with high-resolution images.